caBIG cancer Biomedical Informatics Grid

Meeting/Teleconference Name, Date

Genome Annotation SIG Meeting Minutes

Date, Time & Location:	October 7, 2004 3:00-4:00 EDT
Attendees:	Xiaopeng Bian NCICB Terry Braun University of Iowa Ed Frank University of Chicago Michael Keller Booz Allen Hamilton Juli Klemm 3rd Millennium Rakesh Nagarajan Washington University Harold Riethman Wistar Institute Craig Street Penn Baris Suzek Georgetown University Xiaming Wang University of Chicago
Introduction:	**Roll-call, open meeting, review meeting goals** (Craig Street)
Review Discussions:	**Introduce new SIG lead** (Juli Klemm) Introduced the new SIG lead Craig Street from the Biomedical Informatics Facility at the University of Pennsylvania. **Update on support from the cross-cutting workspace** (Juli Klemm) Representatives from architecture workspace will provide guidance at the project level while representatives from V&CDE workspace will provide guidance at the SIG level. Dan Geraghty from Fred Hutchinson Cancer Research Center will be the V&CDE facilitator for this SIG. Project level pairing for Architecture support will be available next week. **Update on use cases** (Juli Klemm) Received all the use cases for architecture group to look at and will post the use cases next week. Have set up meetings with many of the submitters to review their use cases with the caGRID team. Summary of current findings will be presented at the next ICR WS/WG meeting. Harold wanted to know if we could continue submitting use cases. Answer: Yes. **UML-based CDE-creation training** (Juli Klemm) NCICB and Scenpro have put together a web-based training session for UML-based CDE-creation, including how to annotate a UML model to support CDE creation in the caDSR.
White Paper Discussion:	**White Paper on a 'Gene' Common Data Element** (Rakesh Nagarajan et al.) Rakesh discussed his white paper including a potential list of acceptable gene identifiers. This paper can be found at: http://cabig.nci.nih.gov/workspaces/ICR/Meetings/SIGs/gene_annotation/20041007_GeneCDE_whitepaper Rakesh solicited input from the group to pare this list down to maybe two or three acceptable identifiers. Craig still felt that one common identifier

should be presented and that the others could be listed as aliases. Ed agreed that a small set of identifiers would allow UML models across the grid to be joined automatically provided the same CDE's are used. Harold mentioned that eventually genome annotation will need to incorporate non-coding areas including SNP's, chromosome break-points, etc.

Baris stated that TrEMBL, SwissProt and PIR entries are now merged under UniProt. PIR-PSD will retire by the end of this year. He also suggests to include UniParc (UniProt Archive) ID to the list. Some of the attendees asked about proprietary SwissProt data. All SwissProt data will be public as a part of UniProt project starting next year.

The group discussed the possibility of a caBIG "mapping service" that would resolve gene identifiers. This could eliminate the need to agree upon a common identifier.

The group agreed to convene a smaller subgroup to further discuss the issue of the Gene data element. This group will refine the white paper started by Rakesh and then distribute it to the rest of ICR and to the VCDE workspace for review.

Other discussion items:

caBIG Forum (Craig Street)

Craig and Juli encouraged the group to use the caBIG forum to foster communication within the group.

Action Items:

Name Responsible	Action Item	Date Due	Notes
Juli Klemm	Post the use cases	10/15/04	
Craig Street	Assemble members and schedule CDE Gene "Focus Group"	10/11/04	
Craig Street	Distribute meeting minutes	10/18/04	